



Watson, A. K., Williams, T. A., Williams, B. A. P., Moore, K. A., Hirt, R. P., & Embley, T. M. (2015). Transcriptomic profiling of host-parasite interactions in the microsporidian *Trachipleistophora hominis*. *BMC Genomics*, 16(1), [983]. <https://doi.org/10.1186/s12864-015-1989-z>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1186/s12864-015-1989-z](https://doi.org/10.1186/s12864-015-1989-z)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

RESEARCH ARTICLE

Open Access



Transcriptomic profiling of host-parasite interactions in the microsporidian *Trachipleistophora hominis*

Andrew K. Watson¹, Tom A. Williams^{1*}, Bryony A. P. Williams², Karen A. Moore², Robert P. Hirt¹ and T. Martin Embley¹

Abstract

Background: *Trachipleistophora hominis* was isolated from an HIV/AIDS patient and is a member of a highly successful group of obligate intracellular parasites.

Methods: Here we have investigated the evolution of the parasite and the interplay between host and parasite gene expression using transcriptomics of *T. hominis*-infected rabbit kidney cells.

Results: *T. hominis* has about 30 % more genes than small-genome microsporidians. Highly expressed genes include those involved in growth, replication, defence against oxidative stress, and a large fraction of uncharacterised genes. Chaperones are also highly expressed and may buffer the deleterious effects of the large number of non-synonymous mutations observed in essential *T. hominis* genes. Host expression suggests a general cellular shutdown upon infection, but ATP, amino sugar and nucleotide sugar production appear enhanced, potentially providing the parasite with substrates it cannot make itself. Expression divergence of duplicated genes, including transporters used to acquire host metabolites, demonstrates ongoing functional diversification during microsporidian evolution. We identified overlapping transcription at more than 100 loci in the sparse *T. hominis* genome, demonstrating that this feature is not caused by genome compaction. The detection of additional transposons of insect origin strongly suggests that the natural host for *T. hominis* is an insect.

Conclusions: Our results reveal that the evolution of contemporary microsporidian genomes is highly dynamic and innovative. Moreover, highly expressed *T. hominis* genes of unknown function include a cohort that are shared among all microsporidians, indicating that some strongly conserved features of the biology of these enormously successful parasites remain uncharacterised.

Keywords: Host-parasite interactions, Microsporidia, Gene expression, Transcriptomics, Evolution

Background

Microsporidia are a group of obligate endoparasitic fungi [1, 2]. They are highly successful pathogens that are able to infect a diverse range of hosts, including species of economic significance, and can also cause disease in immunocompromised humans. Microsporidia were first characterised as the causative agent of pébrine, the disease that contributed to the fall of the European silk-worm industry in the 19th century [3]. More recently, the microsporidian *Nosema ceranae* has been implicated

in colony collapse disorder affecting bee populations world wide [4]. Documented cases of human microsporidiosis have risen sharply since the onset of the AIDS pandemic, and 14 species of Microsporidia have been described as causing opportunistic infection in immunocompromised humans [5], including *Trachipleistophora hominis*, the focus of the present study, which was isolated from an HIV/AIDS patient in 1996 [6, 7].

Microsporidians are obligate intracellular pathogens, and their genomes have become highly reduced in terms of protein coding content as a result [8–11]. Recent analyses suggest that most gene loss occurred in the common ancestor of microsporidians, leaving a small core of

* Correspondence: tom.williams2@ncl.ac.uk

¹Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK

Full list of author information is available at the end of the article

genes that carry out functions that are essential for all eukaryotic cells, embellished by an additional conserved core of genes common to most microsporidians [10, 11]. This novel, Microsporidia-specific core includes genes for a variety of different proteins, many of unknown function, and some of which form expanded gene families in contrast to the general trend of microsporidian genome reduction. Included among the latter are genes for nucleotide transport proteins (NTT) that are present in multiple copies on all but one of the microsporidian genomes sequenced to date [8, 10, 11, 12]. In the absence of ATP production capabilities these proteins, which are among the very few proteins to be functionally characterised for microsporidians, play a key role in the import of ATP and other nucleotides from infected host cells, without which the parasite cannot complete its life cycle [13, 14]. The observation that other proteins, including additional transporters, show similar patterns of retention and expansion [11] suggests that they may also play conserved roles in the microsporidian intracellular lifestyle.

The advent of RNA-Seq provides the opportunity to investigate microsporidian gene expression during the intracellular stages of infection on a genome-wide scale. Gene expression analyses of *Encephalitozoon cuniculi* [15], *Nematocida parisii* [16], *Spraguea lophii* [17], and *Nosema bombycis* [18] have already shown that this technology can be used for microsporidians, highlighting the potential of this technique for studying a group of parasites that cannot be genetically manipulated in the laboratory. They have also highlighted strategies by which host cells respond to microsporidian infections, including defence responses mediated by ubiquitination [19], the production of antimicrobial peptides [18], and the perturbation of metabolic pathways [18]. In the present study we have used RNA sequencing to investigate gene expression by *Trachipleistophora hominis* infecting a mammalian (rabbit kidney) cell line, and we compare host expression under infected and non-infected conditions.

Our expression analyses confirm the large coding capacity – 3153 genes – of *T. hominis* refuting a recent suggestion [20] that the large number of genes initially reported might be an artefact of genome annotation. Although our analyses did identify some false gene models, this was balanced by the identification of genes previously missed during the genome annotation, including some that expand the metabolic capacities of *T. hominis* in interesting ways. Parasite gene expression was extremely heterogeneous with 5 % of genes accounting for over 50 % of total gene expression. This includes a strong signature for genes involved in replication and growth, but also a cohort of highly expressed genes that are conserved among microsporidians but are of so far unknown function. We detect strong evidence of functional divergence within gene families including transport proteins that,

counter to the prevailing mode of gene loss, have undergone expansion. These data support classical ideas of functional divergence after gene duplication with the most conserved paralogue being the most highly expressed. Analysis of single nucleotide polymorphisms (SNPs) and their allele frequency spectrum strongly suggest that *T. hominis*, like other unikaryotic Microsporidia [16, 21, 22], is diploid and raises the possibility of a sexual stage in its lifecycle. The detection of active insect-derived transposons suggests that *T. hominis* – which is an opportunistic pathogen of humans – has an insect host in nature. Transcriptional profiling of the host is consistent with a generalised cellular shutdown upon infection with *T. hominis*, with the upregulation of host pathways for ATP production, amino sugar and nucleotide sugar metabolism: these pathways potentially complement gaps in parasite biosynthesis predicted by *in silico* analyses.

Results and discussion

Reproducibility of host-parasite transcriptomics

T. hominis is an obligate intracellular parasite grown in laboratory co-culture within rabbit kidney (RK) cells [6]. We harvested total RNA from three biological replicates of infected RK cells seven days post inoculation, at which point ~60 % of RK cells in each flask were infected with *T. hominis*. At this stage the community of *T. hominis* cells was a mixture of different life cycle stages including thick walled spores and pre-spore stages (sporonts and sporoblasts) as well as the intracellular sporoplasm (newly germinated parasite inside the host cell) and replicative or meront stage (Fig. 1). Although we used a bead beating method similar to one previously shown [10] to lyse *T. hominis* spores in our RNA extractions, the resistant nature of the spore-forming stages of the parasite lifecycle may make lysis less efficient, which could lead to an enrichment of transcripts from replicative stages in the total RNA pool; possible implications of this bias are discussed in more detail below. In parallel we also isolated total RNA from three biological replicates of uninfected RK cells, in order to compare patterns of host expression under the two conditions. We obtained 2.3×10^7 sequencing reads from the infected cells, with 7.7 % of these reads mapping to the *T. hominis* genome. The reproducibility between biological and technical replicates was very high, both for pairwise comparisons of the expression of individual genes between replicates and the overall distribution of expression levels across all transcripts (Fig. 2 (*T. hominis*); Additional file 1: Figure S1 (rabbit)). These results indicate that our analysis of the *T. hominis* and host transcriptomes was highly reproducible; thus, the potential biological implications can be explored in a meaningful way.

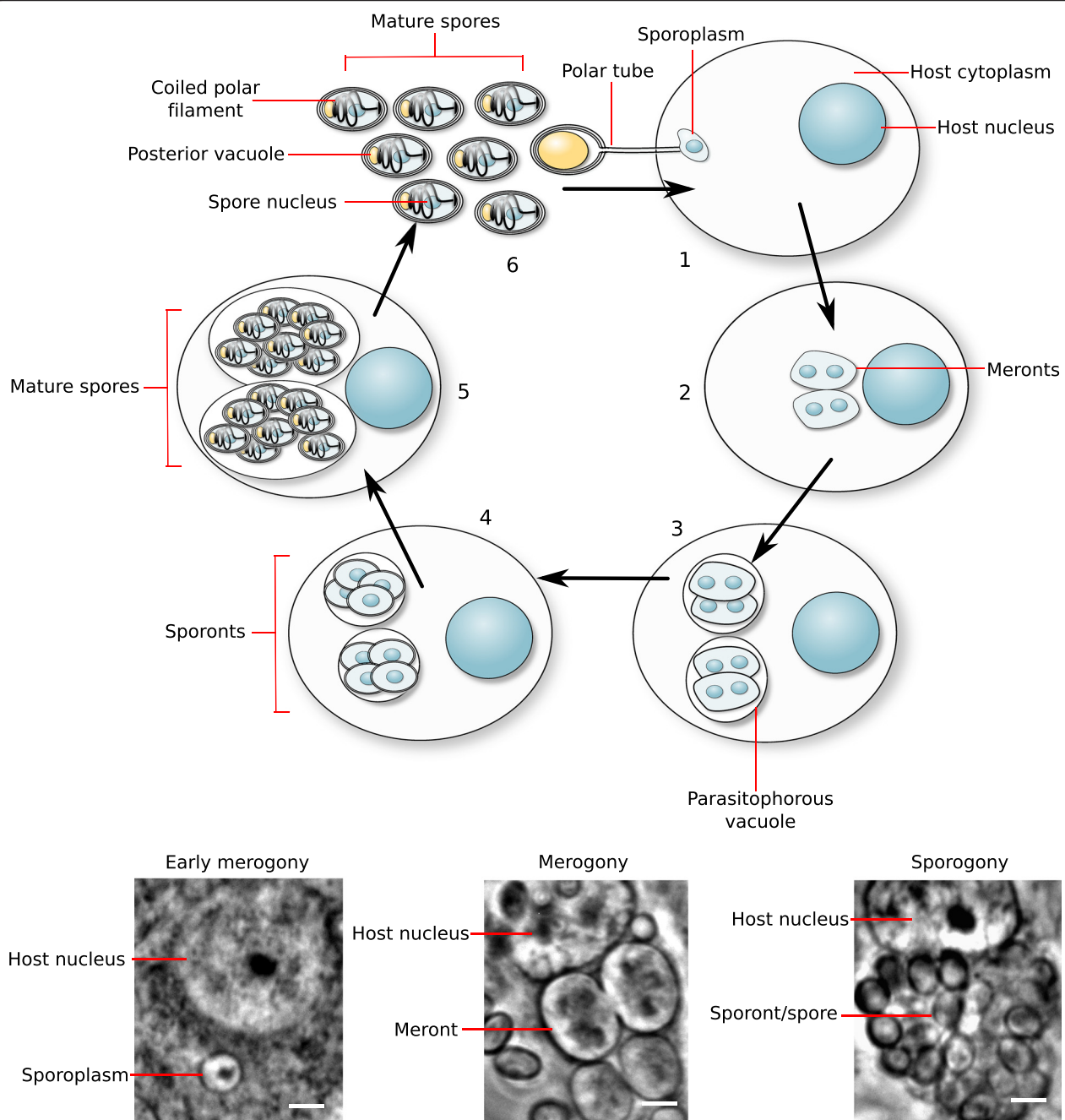


Fig. 1 A schematic lifecycle of *Trachipleistophora hominis* with phase contrast images of representative intracellular stages. In stage 1, the *T. hominis* spore injects its contents into the host cell via a proteinaceous polar tube. This uninucleate sporoplasm is the earliest stage of infection of the host cell [119]. Following injection, the parasite – now called the meront – grows and begins to proliferate in the host cell (merogony, the second stage of the parasite lifecycle) [6, 7]. It is unknown how many cycles of proliferation meronts undergo in the host cell before stage three of the lifecycle begins, the transition from meronts to spore-forming sporonts. In Microsporidia this is marked by the deposition of an electron-dense material on the surface of the parasite, and in *T. hominis* may represent the timing of formation of the parasitophorous vacuole, an additional membrane surrounding the parasite within the host cell [6, 7, 119]. During spore formation (sporogony) the microsporidian spore wall is assembled, and *T. hominis* undergoes a further round of proliferation [6, 7]. Eventually mature spores form inside the host cell (stage 5). These spores may either eject their polar tubes from within their host cell and infect neighbouring cells, or else be released from the cell in stage 6 of the parasite lifecycle, allowing dissemination of the infection to new cells or new organisms [119]

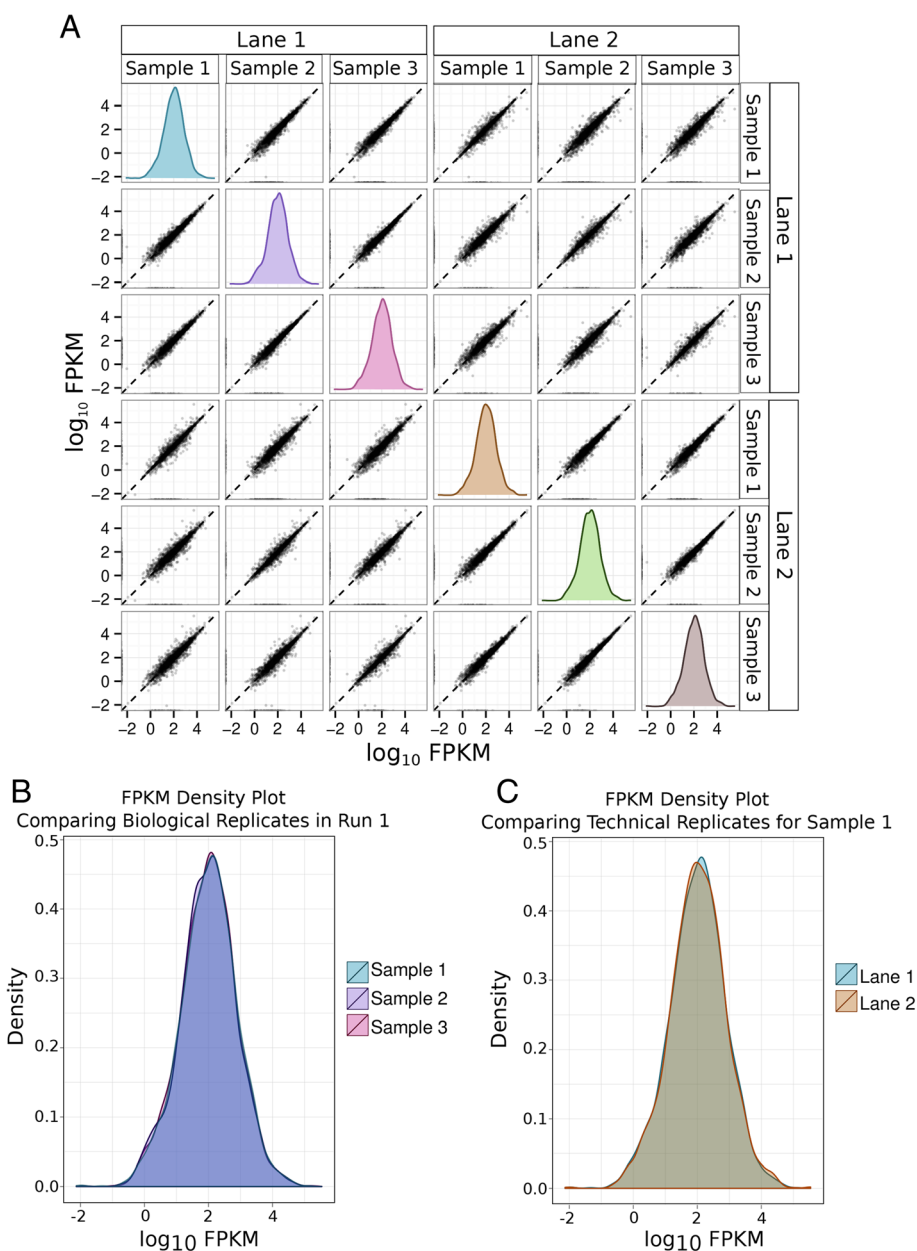


Fig. 2 Robustness and reproducibility of transcriptomic analysis for *T. hominis*. In this figure, each sample is an individual biological replicate of *T. hominis* infected RK-13 cells from which total RNA was purified. All of these samples were sequenced on two different lanes of the Illumina sequencing chip to provide technical replication. The strong positive correlation of transcript abundance, as measured by Fragments Per Kilobase per Million mapped reads (FPKM) between replicates in all cases is indicative of high levels of reproducibility between both biological and technical replicates. **a** Pairwise comparisons of *T. hominis* transcript \log_{10} FPKM values between biological replicates (samples) and technical replicates (sequencing lanes); density plots generated in cummeRbund [109] represent the distribution of \log_{10} FPKM values for that replicate. **b** FPKM density plot overlay comparing the distribution of *T. hominis* transcript FPKM values between individual biological replicates (samples) from a single sequencing lane. **c** FPKM density plot overlay comparing the distribution of *T. hominis* transcript FPKM values between technical replicates (different sequencing lanes) from a single sample (Sample 1). The levels of reproducibility between replicates for the expression of host genes were equally very high, this data is summarised in Additional file 1, Figure S1

***T. hominis* has more genes than small-genome microsporidians**
The *T. hominis* genome (~11.6 Mbp haploid genome size) was initially reported to contain 3266 predicted

open reading frames (ORFs), which is over 1000 more genes than the ~2000 open reading frames predicted for the best studied small genome (~2.3–2.5 Mbp haploid genome size) species of *Encephalitozoon* [20]. It has been

suggested that the predicted number of *T. hominis* genes may be an overestimate due to over-prediction of small genes [20]. Indeed, some ORFs (113) appear to be unique to *T. hominis*, with no homologues identified in its closest sequenced relative *Vavraia culicis* [23, 24], or any other microsporidian [10, 11]. Additionally, *T. hominis* encodes a large family of novel leucine rich repeat proteins (117 ORFs) that includes many fragmented ORFs, indicative of ongoing pseudogenisation [10]. It is therefore possible that the estimated coding capacity of *T. hominis* includes some false positives, particularly given the highly derived nature of most microsporidian gene sequences [10, 20]. We obtained evidence for the expression of 2958 (90 %) of the 3266 annotated *T. hominis* ORFs (Additional file 2: Table S1), including 85 % of the leucine-rich repeat genes, suggesting that most are genuine ORFs. However, we detected expression for fewer (73 %: 83 of 113) *T. hominis*-specific genes, consistent with the idea that some of these are false-positive calls. Balancing this reduction in the predicted gene count based on transcription, we obtained evidence for an additional 292 transcripts that were not predicted as ORFs in the original genome project. One hundred and fifty-five (80 %) of these transcripts are located within regions of ambiguous genomic sequence or near the ends of scaffolds; the difficulty in annotating these regions may explain their absence from the original *T. hominis* genome annotation [10].

Ninety seven of the 292 novel transcripts lacked an ORF suggesting that they might be *T. hominis* noncoding RNAs, although they did not give significant matches to the noncoding RNAs already included in the Rfam database [25, 26]. They also appear to be missing from other microsporidian genomes as searched using BLASTN. Despite this, one of the transcripts, XLOC_000764, was in the upper 95th percentile of overall expression levels; that is, its expression level was higher than 95 % of detected transcripts, suggesting that it plays a physiologically relevant role. The remaining 195 transcripts are predicted to contain ORFs of which 89 had significant hits to the nr protein database at a BLASTX E-value cutoff of 0.01. The two most highly expressed of the 89 shared significant similarity to partial ORFs (VCUG_01670 and VCUG_016701) annotated in the *Vavraia culicis* genome [24], the closest sequenced relative of *T. hominis* [23]. The putative *T. hominis* protein is 564 amino acids in length and contains a series of 12 tandem glycine-asparagine repeats. A search using HHPred [27, 28] suggests that these are similar to a class of repeats present in over 30 % of *Plasmodium falciparum* genes. Their function in *Plasmodium falciparum* is unknown, but it has been suggested that they may interact with host proteins [29]. Consistent with this idea, the *T. hominis* gene has an N-terminal signal peptide [30], suggesting

that it might be secreted or localised on the surface of the parasite.

Twenty-five of the newly identified transcripts show significant similarity to genes outside of the microsporidian clade, including several broadly-distributed eukaryotic genes previously thought to be absent from the genome of *T. hominis*. These include exportin, a component of the nuclear export machinery, as well as homologues of deoxyhypusine hydroxylase [31], the Rea1 AAA-ATPase [32], and the 60S ribosomal protein L29 [33]. We found transcript evidence for three new transport proteins, including an amino acid/auxin permease of the AAP family, a putative cation transporting P-type ATPase, and a member of the DMT superfamily of drug and metabolite transporters that includes a UAA transporter family domain (pfam:08449) associated with UDP-N-acetyl-glucosamine:UMP antiporter activity. Published data demonstrate that microsporidians can import purine nucleotides using nucleotide transport (NTT) proteins [11, 13, 14] but as yet there is no evidence for the transport of pyrimidine nucleotides by these transporters. Based upon *in silico* predictions it appears that microsporidians cannot make pyrimidines *de novo* so there is a transport gap that needs to be filled [11, 13, 14]. UDP-N-acetylglucosamine is the direct monomeric precursor for chitin synthesis, an integral component of the microsporidian spore wall [34], but it is also biosynthesised by the mammalian host cell in which it plays roles as a co-enzyme, signalling molecule, and precursor for glycosylation [35, 36]. The pyrimidine UDP is liberated from UDP-N-acetylglucosamine during chitin polymerisation by chitin synthase and during glycosylation, so as well as providing chitin precursors this novel transporter could potentially provide the starting substrate to make pyrimidines needed for *T. hominis* DNA and RNA biosynthesis (Fig. 3). Genes within the chitin biosynthesis pathway in *T. hominis* are generally expressed at similar levels in our analysis (600–900 fpkm), however expression of the terminal components of the pathway, chitinase and chitin synthases, were much lower (20–30 fpkm) (Fig. 4). These observations suggest that in the proliferative stages of the microsporidian lifecycle, UDP-N-acetylglucosamine is either being used primarily for UDP-liberating glycosylation reactions, or being accumulated for chitin production during sporogony. We also obtained transcriptomic evidence for a previously unannotated *T. hominis* homologue of a UDP-N-acetylglucosamine pyrophosphorylase, an enzyme also encoded on the genomes of *Vavraia*, *Encephalitozoon*, *Anncalia*, *Edhazardia* and *Vittaforma*. This enzyme catalyses the conversion of UTP and N-acetyl-alpha-D-glucosamine 1-phosphate to UDP-N-acetylglucosamine and diphosphate. This means that, in addition to potential acquisition of UDP-N-acetylglucosamine from the host, *T. hominis* encodes a complete pathway for its biosynthesis

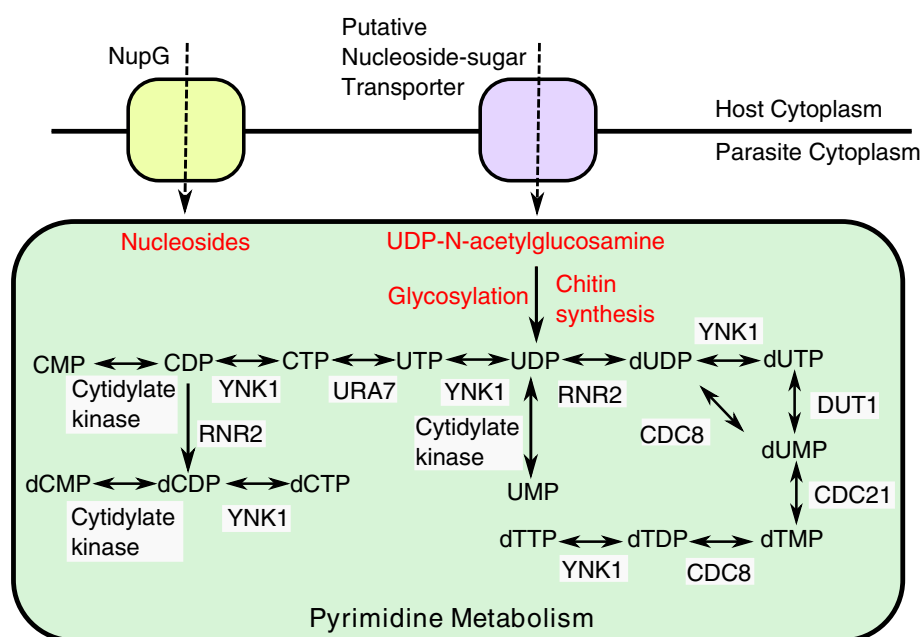


Fig. 3 A potential route for pyrimidine and chitin precursor acquisition by *T. hominis*. *De novo* routes for pyrimidine biosynthesis are not present in Microsporidia, but scavenger pathways allow conversion between different pyrimidines [11]. The import of only one pyrimidine from the host cell would, in principle, enable the biosynthesis of other pyrimidines via scavenger pathways, because interconversions between bases are possible. NupG-like transport proteins have been hypothesised to transport nucleosides and thus act as a source of pyrimidine precursors [16], but the specificity and subcellular localisation of these genes are currently unknown. A putative UDP-N-acetyl glucosamine transporter provides another potential source of pyrimidines in *T. hominis* and *V. culicis*: the imported metabolite can feed into chitin synthase and glycosylation pathways, both of which may liberate free pyrimidine

(Fig. 4). The potential importance of this enzyme for spore wall formation in Microsporidia makes it a potential target for therapeutic intervention. Thus, recent studies have shown that UDP-N-acetylglucosamine pyrophosphorylase is essential for the survival of *Trypanosoma brucei* in its bloodform lifecycle stage [37], and chemicals that can selectively inhibit the *Trypanosoma brucei* UDP-N-acetyl pyrophosphorylase have been identified [38].

In conclusion, the transcript data from our experiments appear to be very reproducible for both host and *T. hominis*. The data validate the majority of gene models originally predicted from the *T. hominis* genome sequence and identify some previously missed genes, providing evidence for a total of 3153 transcribed genes and making the *T. hominis* gene complement one of the largest identified for microsporidians so far [20].

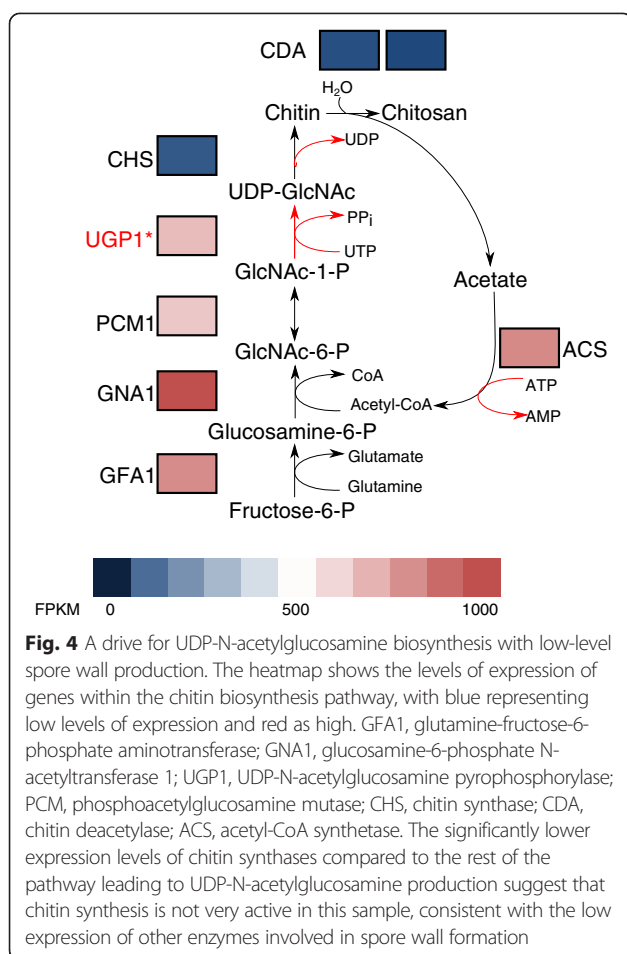
Overlapping transcription in a gene-sparse microsporidian genome

We detected 140 *T. hominis* transcripts encoding more than one ORF, of which 113 overlap on the genome assembly. The remaining 27 do not overlap on the genome, but the intergenic region between them was spanned by RNA-Seq reads. This suggests that

overlapping transcription occurs in *T. hominis*, as previously reported for the small genome species *E. cuniculi* and *A. locustae*, where a similar number of multi-ORF transcripts (144) was identified [39, 40]. Interestingly, our data suggests that overlapping transcription is not necessarily linked [40, 41] with genome compaction, because gene density in *T. hominis* is actually lower than for yeast [10]. One possibility is that overlapping transcription provides a mechanism for co-regulating the expression of particular genes [42]. If so, then this mode of regulation is poorly conserved over evolutionary time, because transcriptional overlap is conserved for only one gene pair across *E. cuniculi*, *A. locustae*, and *T. hominis* [40]: ribosomal protein L6 (THOM_0162) and RNA polymerase III transcription factor IIIC subunit 5 (THOM_0163), which overlap in the 3' UTR.

Low levels of splicing in *T. hominis*

A total of 85 introns were predicted in *T. hominis* based upon the presence of conserved *E. cuniculi* intron motifs [10]. Surprisingly, we obtained transcriptomic evidence for the splicing of only a single gene – 40S ribosomal protein S23 – at this conserved intron motif (JUNC000000002 in Additional file 3: Table S2); this gene



was spliced efficiently, with 93 % of detected transcripts being spliced, higher than the splicing efficiency detected for any spliced gene in the *E. cuniculi* transcriptome [15]. This gene was also one of only two genes for which splicing was detected in transcripts from the microsporidian *Spraguea lophii*, and the intron has a similar length [17]. Splicing in general in Microsporidia appears to be inefficient for short introns [17], with splicing rates of less than 15 % identified for several *E. cuniculi* transcripts [15]. This potentially explains our inability to detect splicing at the great majority of characterised intron motifs in the *T. hominis* transcriptome. By contrast, we obtained evidence for 13 introns where RNA sequencing reads suggest a spliced exon boundary. One of these novel junctions (JUNC00000048 in Additional file 3: Table S2) is found in the third most highly expressed transcript in our dataset, and its presence is supported by a greater number of reads than the originally predicted 40S ribosomal protein S23. Surprisingly, the remaining new experimentally identified introns lack the classical intron motif previously described in *E. cuniculi* [43], and an alignment revealed no unique novel motif common to the group (Additional file 4: Figure S2).

Highly expressed *T. hominis* genes

The top 5 % (170 genes) of genes accounted for over half (58 %) of all detected transcription in *T. hominis* (Additional file 5: Figure S3). Fifty three percent of these highly transcribed ORFs belong to a core conserved set of microsporidian genes defined by Nakjang et al. [11] as those encoded by 9 of the analysed 11 sequenced microsporidian genomes. Of these highly expressed core genes 32 % encode rRNA or ribosomal proteins, and 31 % encode other essential elements of eukaryotic cell biology including cytoskeletal proteins, transcription and translation factors, cell division proteins (e.g., NudC), histones, and molecular chaperones. Similar functional gene groups (E.g. Ribosome biogenesis and protein translation factors) were significantly enriched in highly expressed genes during proliferative growth of fission yeast (*Schizosaccharomyces pombe*) [44], suggesting the pattern of expression observed in *T. hominis* may reflect the signatures of its rapid growth and replication in the host cell. The highly expressed core microsporidian genes also include a large number of microsporidian hypothetical proteins of unknown function (37 %). In fission yeast, groups of functionally related genes tend to be expressed at similar levels [44], suggesting that if the same pattern applies to Microsporidia these uncharacterised proteins may also play important and as yet unknown roles in core parasite biology and proliferation.

The inclusion of molecular chaperones of the HSP40/DNAJ and HSP70 families and protein disulfide isomerase among highly expressed core *T. hominis* genes is intriguing. High expression of HSP70 has been identified in several different microsporidian clades [41], suggesting it is a common feature of the group. Chaperones have been identified as important virulence factors in bacterial and eukaryotic pathogens, including intracellular bacteria [45]. Published work has demonstrated that intracellular bacteria and bacteria cultured under conditions that introduce a population bottleneck often over-express chaperones to maintain functionality under an increased mutational load [46–50]. The high levels of chaperonin expression in several microsporidians, which also experience bottlenecks during transmission and show high rates of sequence evolution [6, 10], suggest that these intracellular eukaryotes are behaving in the same way [41].

Replication and biosynthesis are energy-requiring processes and hence *T. hominis* must either make or acquire ATP and GTP during proliferation. Genome analyses suggest that *T. hominis* has a complete glycolytic pathway [10] and we detected expression of all of the relevant enzymes in the transcriptome, but only glyceraldehyde 3-phosphate dehydrogenase was in the top 5 % of transcripts. The most highly expressed glycolytic enzymes are PBA, GAPDH and PGK, with GAPDH

and PGK respectively providing NAD^+ reduction and ATP synthesis. These data suggest that *T. hominis* is potentially making some of the ATP it needs, but the cell stage where this occurs is not resolved by our data. Quantitative immuno-localisation of PGK protein, the first ATP-generating step of glycolysis, suggests that the protein is mainly, but not exclusively, inside *T. hominis* spores rather than vegetative cells [10]. It has also been previously suggested that glycolysis occurs mainly in the spores of another microsporidian, *Paranosema grylli* [51]. Interestingly, transcripts for the most abundant protein identified in spores, Polar Tube Protein 3, were detected at only modest levels in the RNA sequencing data, with an average of $18.1 \text{ FPKM} \pm 12.3 \text{ SD}$ (within the 30th percentile for expressed genes in the dataset). Indeed, no genes known to be associated with spore wall formation (spore wall proteins, polar tube proteins and chitin synthases) were found in the top 5 % of highly transcribed genes. One possibility is that the lysis procedure used to extract total RNA may not lyse *T. hominis* sporonts, sporoblasts and spores very efficiently; if so, then the extracted RNA would be enriched for transcripts from replicative stages of the parasite lifecycle despite the mixed infection. Alternatively, these more quiescent stages of the parasite lifecycle might show an overall reduction in levels of gene expression, as observed in the nonreplicating stages of the fission yeast cell cycle [44], so that they will naturally be represented at lower levels in the total RNA pool. Conversely, high levels of expression were observed for a number of genes involved in DNA replication and proliferation, consistent with an enrichment of transcripts from the actively proliferating stages of the parasite, and suggesting that some ATP production by glycolysis may occur in these stages of the parasite lifecycle.

Few metabolic enzymes appeared in the top 5 % of expressed transcripts, consistent with genomic predictions that *T. hominis* must import many of the substrates it needs for biosynthesis directly from the infected host cell [10, 11]. One highly expressed enzyme is nucleotide diphosphate kinase (YNK) [52], which was also highly abundant in proteomic analyses of highly purified spores [10], YNK is predicted to play a key role in supporting *T. hominis* intracellular proliferation by converting nucleotides or nucleoside diphosphates to their triphosphate forms, the precursors for both RNA and DNA synthesis and sources of cellular energy. High levels of expression were also observed for dUTPase, another enzyme predicted to be involved in nucleotide biosynthesis [53]. The most highly transcribed metabolic enzyme was an asparagine synthetase A (asnA, THOM_2136, InterPro ID: IPR004618). Among Microsporidia, coding sequences for this protein are found on the genomes of *T. hominis*, *V. culicis*, *Enterocytozoon*

bienieusi, *N. ceranae* and *Nosema pernyi*, and were likely acquired by lateral gene transfer from bacteria [10, 11]. Among eukaryotes, AsnA is almost exclusively found in parasites [10] including *Typanosoma brucei* and *Leishmania donovani*, where it is essential for survival [54, 55]. In bacteria, AsnA is responsible for the reversible transamination of aspartate to asparagine in the presence of ATP and ammonia. *Leishmania* and *Trypanosoma* AsnA, the only characterised eukaryotic homologues, have a broader specificity, and are able to use glutamine as a nitrogen donor to generate glutamate [55, 56]. One possibility is that microsporidian AsnA may play important roles in interconversion between essential amino acids, including glutamine, an important precursor to both chitin biosynthesis for spore wall formation and glutathione (GSH) biosynthesis required in parasite detoxification systems. This ORF contains an aminoacyl-tRNA synthetase (class II) domain (InterPro ID: IPR004364) suggesting that it may also function to add asparagine to its cognate tRNA. Its specificity to parasitic eukaryotes, coupled with its functional importance, high expression level, and the availability of an AsnA crystal structure make the protein a promising potential drug target [55].

Maintaining supplies of glutamine is important for the generation of GSH, a detoxifying molecule required for the prevention of damage by reactive oxygen species [57]. The *T. hominis* detoxification system also includes thioredoxin reductases, peroxidases, glutathione reductases and superoxide dismutase [10]. While all identified components of this pathway are expressed in *T. hominis*, the highest expressed in our dataset, and only component in the 95th percentile of overall expression levels, was iron/manganese superoxide dismutase (SOD), which reduces and detoxifies superoxide molecules (O_2^-) [10]. The largest biological source of superoxide species is as a by-product in the production of ATP by oxidative phosphorylation, which *T. hominis* does not carry out but which we show is upregulated in host cells during infection by *T. hominis* (see below). *T. hominis* SOD may protect against oxidative stress generated by the host cell as it supports both its own survival and parasite proliferation. Previous work [58–60] has indicated that oxidative stress in the host cell is elevated during infection; therefore, a robust detoxification system is likely to be important for parasite survival, as observed in some bacterial infections [61].

One metabolic pathway known to be essential and that has been functionally characterised in Microsporidia is iron-sulphur cluster biogenesis [62]. Iron-sulphur clusters are required for the activity of key proteins needed for microsporidian replication, including DNA polymerase. The metabolic pathway for the biogenesis of iron-sulphur clusters is compartmentalised, starting in the

microsporidian mitosome (remnant mitochondrion) and ending in the cytosol [62]. This allows us to examine the variability in levels of transcript abundance between different compartments in a single linked pathway that should be required throughout the parasite lifecycle. The only highly expressed gene in the pathway (in the top 5 % of expressed transcripts) was Dre2, a cytosolic component of iron-sulphur cluster biogenesis [63]. Interestingly, Dre2 also plays a role in inhibiting free radical-induced apoptosis [63]; given that other detoxifying enzymes are also highly expressed in *T. hominis*, it may be this function of Dre2 that drives its high expression level. Consistent with this idea, the other components of iron-sulphur cluster biogenesis are expressed at significantly lower and similar levels, suggesting that genes in the same pathway may be generally expressed at similar levels in *T. hominis*, as observed in fission yeast [44].

Surface-located transport proteins are predicted by genome analyses to be fundamental for supporting the replication of *T. hominis* and other microsporidians by importing substrates from infected host cells [3, 10, 11, 13, 14]. The expression of *T. hominis* proteins related to known transporters, or annotated as potential transporters, is very heterogeneous within structural types. Only three predicted transport proteins are found in the top 5 % (>2210 FPKM) of expressed genes and these do not include any of the *T. hominis* nucleotide (NTT1-4) transporters for which functional data is available [14]. The most highly expressed of these – NTT4 – appears in the 92nd percentile of expression levels. NTT4 is one of four paralogous nucleotide transporters that are expressed on the surface of replicating parasites [14], where they function to transport purine nucleotides including ATP and GTP, for energy and/or biosynthesis [14]. Of the three highly expressed transporters, the first and third in terms of expression are hypothetical transporters of unknown specificity while the second is a putative inorganic phosphate transport protein. Given the potential importance of the NTT transporters it seems possible that these more highly transcribed transporters also support important, albeit currently uncharacterised, cellular functions.

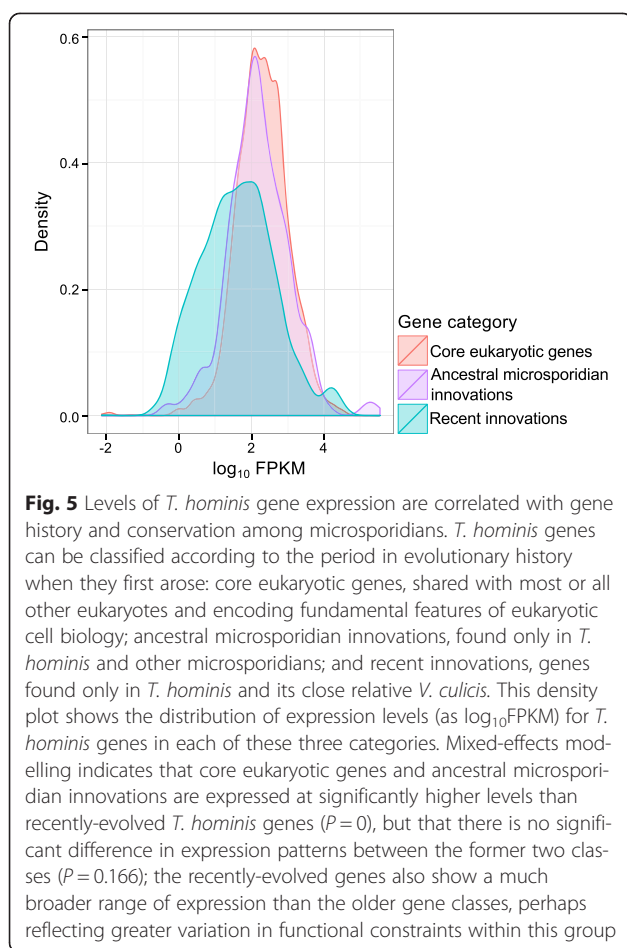
The most highly transcribed of the three membrane proteins (THOM_1886) appears to be specific to *T. hominis* and its closest sequenced relative, *V. culicis*, as determined by sensitive PSI-BLAST [64] and HMMER [65] based searches. The THOM_1886 protein is predicted to include 7 transmembrane domains and was annotated as a putative transport protein [10]. Its expression level is more than 1500x that of the average transporter in our study (22209 FPKM \pm 5483 SD) suggesting that, in addition to the conserved common core of microsporidian genes, lineage-specific innovation is also important for parasite biology.

Levels of *T. hominis* gene expression are correlated with gene history and conservation among microsporidians

Comparative analyses of the genome of *T. hominis* with other microsporidian genomes have demonstrated that genome evolution has been a dynamic process, in which the loss of ancestral gene families has been partially offset by the gain of new microsporidia-specific genes [8–11, 16, 17, 66]. In this broader evolutionary context, *T. hominis* genes can be classified into three major groups: core eukaryotic genes – that is, core microsporidian genes defined in Nakjang et al. [11] that were also found in most or all eukaryotes; ancestral microsporidian innovations, or core microsporidian genes that evolved in the common ancestor of all microsporidia (and thus are not identified in other eukaryotes); and recent innovations (for example THOM_1886) that are only found in *T. hominis*, or that are shared between *T. hominis* and its close relative *V. culicis*. To evaluate the relationship between evolutionary conservation and expression level in *T. hominis*, we compared the expression levels of the genes in these three classes using a linear mixed-effects model (Fig. 5).

Our analysis indicated that core eukaryotic genes and ancestral microsporidian innovations were both expressed at significantly higher levels than recently-evolved genes specific to the *T. hominis/V. culicis* lineage ($P = 0$), but that there was no significant difference in expression levels between the two more highly-expressed classes ($P = 0.166$). The consistently high expression levels for core eukaryotic genes are not, in themselves, particularly surprising: this category includes genes involved in basic cellular processes such as DNA replication and repair, mitochondrial iron-sulphur cluster assembly, intracellular trafficking and in some metabolic pathways such as glycolysis and the pentose phosphate pathway. However, the equally high level of expression observed for ancestral microsporidian innovations is interesting because it implies that genes which first evolved in the common ancestor of microsporidia, and which were then conserved across the group, are as important to microsporidians – by the measure of transcript abundance – as genes encoding the fundamental eukaryotic cellular componentry.

By contrast, genes specific to the *T. hominis/V. culicis* lineage are expressed at significantly lower levels than other genes in the parasite ($P = 0$); we focused on genes shared between these close relatives to minimise the impact of false ORF calls on our analyses. Most of these genes are expressed (735 out of 862, or 85 %), but not necessarily at high levels; as can be seen from Fig. 5, this class of recently evolved genes displays a broad range of expression levels. The more heterogeneous distribution of expression levels for recently evolved genes is consistent with a recently proposed model [67] for the gradual



emergence of proto-genes from previously non-coding sequence. Under this model, some new, fortuitously expressed genes acquire important functions and are maintained by selection, while others do not and will eventually be lost to drift and pseudogenisation. It is possible that this process of genomic innovation underpins recently evolved host-parasite interactions for these two species, both of which are thought to infect insects as their natural hosts [10, 68]. Consistent with this hypothesis, the *T. hominis*/*V. culicis*-specific families are enriched for signal peptides ($P=1.2 \times 10^{-10}$, Fisher's exact test) [11], suggesting that the proteins in these families may be localised to the parasite cell surface, part of the infective polar tube, or secreted into the host cell.

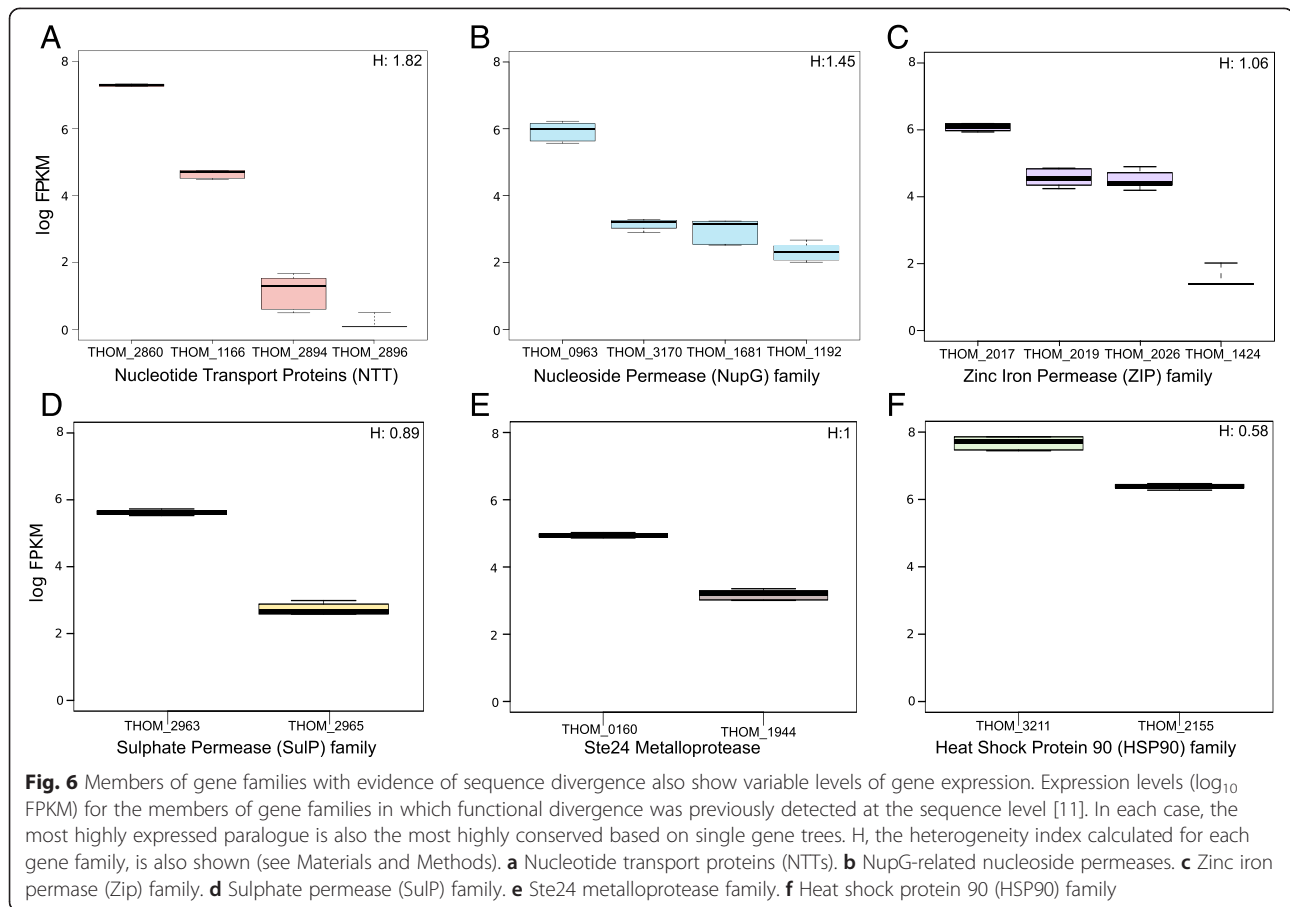
Expression divergence in expanded *T. hominis* gene families

In contrast to the general trend of reductive evolution among microsporidians, a number of *T. hominis* gene families have expanded through gene duplication. Gene duplication is important in the evolution of gene family function, because duplication events can relax selective constraints allowing the functions of one or both

paralogues to change [69, 70]. Consistent with a role for duplication and functional divergence in microsporidian evolution, Nakjang et al. [11] and Heinz et al. [14] found evidence of sequence divergence at conserved amino acid residues following microsporidia-specific duplications in the Hsp90 chaperone, Ste24 metalloprotease, NTT nucleotide transporter, ZiP zinc ion permease, SulP sulphate permease and NupG nucleoside transporter families. Intriguingly, members of these expanded gene families tend to be expressed at above-average levels in *T. hominis* ($P=3 \times 10^{-4}$, linear mixed-effects model).

Figure 6 summarises expression levels for the functionally characterised *T. hominis* nucleotide (NTT) transport proteins [14] and *T. hominis* members of the microsporidian gene families investigated by Nakjang et al. [11]. The variation in expression level is clearly correlated with the evolutionary history of the gene family: in all of these cases, the most highly conserved family member, in terms of conservation of critical residues or branch length in gene family trees [11, 14], is also the most highly expressed. These data are consistent with a model of functional divergence whereby one conserved, highly-expressed paralogue continues to carry out the ancestral function, while other duplicates experiencing reduced selective constraint can gain new functions [71–73]. Any new functions, which could include stage-specific expression, different cellular location or substrate affinity or specificity, will need to be identified through experiment. For example, proteomics data already suggest that NTT4, the most highly expressed member of the gene family (Fig. 6), is the main NTT transporter located within the *T. hominis* spore [10]. In *Encephalitozoon cuniculi* the NTT transporter family has undergone an independent expansion [14]. This expansion was followed by divergence in both sequence and localisation, with one family member (EcNTT3) localised to the mitosome while the other three are located on the surface of replicating parasites [13]. The correlation between sequence divergence and transcript abundance we observe in *T. hominis* is maintained in *E. cuniculi* [15], with divergent family members (EcNTT3 and EcNTT4) expressed at lower levels (mean FPKM 139 for EcNTT3, 96 for EcNTT4) than the more highly conserved family members (EcNTT1 and EcNTT2 – 402 and 977 FPKM respectively).

To evaluate variation in expression among microsporidian gene duplicates more systematically, we analysed expression for all duplicate families identified by Nakjang et al. [11] containing at least two paralogues in *T. hominis*. We calculated the standard deviation of FPKM values within each family, and normalised by the per-family mean (see Materials and Methods) (Additional file 6: Table S3). Expanded *T. hominis* families were then ranked by this metric to identify the families showing



the most extreme expression divergence. Plotting these scores revealed an inflection point in the distribution of the metric, above which we considered within-family expression to be highly heterogeneous (Additional file 7: Figure S4). The most heterogeneous family identified by this approach included a family of retrotransposon-encoded reverse transcriptases; some family members had no detectable expression, suggesting ongoing pseudogenization as might be expected for transposable elements. The group of *T. hominis* families showing the greatest expression divergence also includes the hexokinase gene family, whose paralogues in the microsporidian *N. parisii* have been suggested to manipulate host metabolism following secretion into the host cell [16]. *T. hominis* encodes four hexokinases of which two include predicted signal peptides [11], consistent with the hypothesis that they may be secreted into the host cell. One of the remaining copies (XLOC_001491) is the most highly expressed member of the family, again consistent with the idea that the most highly conserved member of a duplicated family continues to perform the ancestral function.

Parallel horizontal transfers of transposons implicate an insect host in the lifecycle of *T. hominis*

T. hominis is one of several microsporidians that retain elements of the RNA interference machinery [10]. The core components of this machinery, Dicer and Argonaute, are both expressed by *T. hominis* during infection but are not in the top 5 % of expression. The RNAi machinery is hypothesised to play a role in defence against transposon activity in Microsporidia [10]. Consistent with this hypothesis, we obtained evidence for the expression of 58 of the 110 annotated transposons in the *T. hominis* genome. Combined with evidence for transcription of transposons in *Edhazardia aedis* [41], these data suggest that active transposons pose an ongoing threat to genome integrity in microsporidians more generally.

Although *T. hominis* is an opportunistic parasite of immunocompromised humans, its natural host remains unknown. *T. hominis* can proliferate within artificially infected mosquitoes under experimental conditions, but these infections have not been observed in nature [74]. One of the novel transcripts identified in this study showed similarity to a PiggyBac transposase [75]. This

transcript maps to a previously unannotated portion of the *T. hominis* genome, and is therefore distinct from the PiggyBac element reported in the original genome annotation [10]. The best BLAST hits to the novel PiggyBac element include transposable elements from insects but not other Microsporidia, raising the possibility that *T. hominis* gained the element in a recent horizontal transfer that occurred after the divergence of *T. hominis* from its close relative *V. culicis*. A PiggyBac element identified as recently acquired in bats has been demonstrated to retain activity when inserted into both human and yeast cells, highlighting the capacity of this particular family of transposons for inter-species transfer [76]. Phylogenetic analysis of the novel *T. hominis* element strongly suggests that it was recently acquired from an insect, and probably a member of the Hymenoptera (ants, bees and wasps; Fig. 7 and Additional file 8: Figure S5). The *T. hominis* sequence forms a strongly supported clade with PiggyBac elements from bees (*Bombus impatiens* and *Megachile rotundata*) and the ant *Harpegnathos saltator* (Fig. 7, Clade B). Interestingly, this transfer appears to have occurred independently of the previously identified PiggyBac acquisition from insects in *T. hominis*, which branches in a separate insect clade with maximal posterior support (1.0 posterior probability; Fig. 7, Clade A). In both cases, the most closely related sequence

is from Jerdon's jumping ant (*Harpegnathos saltator*), although the posterior support for this relationship is variable (0.99 in Clade A, and 0.77 in Clade B). The phylogeny of the two separate *T. hominis* PiggyBac elements provides consistent support for the hypothesis that *T. hominis* infections of humans may represent opportunistic zoonoses from a natural hymenopteran host. Interestingly, we also identified a separate novel horizontal transfer from insects into the microsporidian *Nosema apis*, a honeybee parasite [77], providing further support for the horizontal transfer of host-derived transposable elements into Microsporidia (Fig. 7, Clade C) [78–80].

Patterns of single nucleotide polymorphism reveal that *T. hominis* is diploid

Microsporidia replicate inside their host cell, and can only exist outside it as a resistant and infectious spore [3]. This would be a barrier to sex between two different parasite cells, which would require multiple independent infections of a single host cell. Nonetheless, limited evidence for a sexual reproduction cycle has been identified in several microsporidians [81]. The early morphological characterisation of *Ambylospora* identified a meiosis-like stage of division including karyogamy, the fusion of two haploid nuclei to form a single diploid nucleus [82]. Recent studies of the *Nosema/Vairimorpha* lineage have

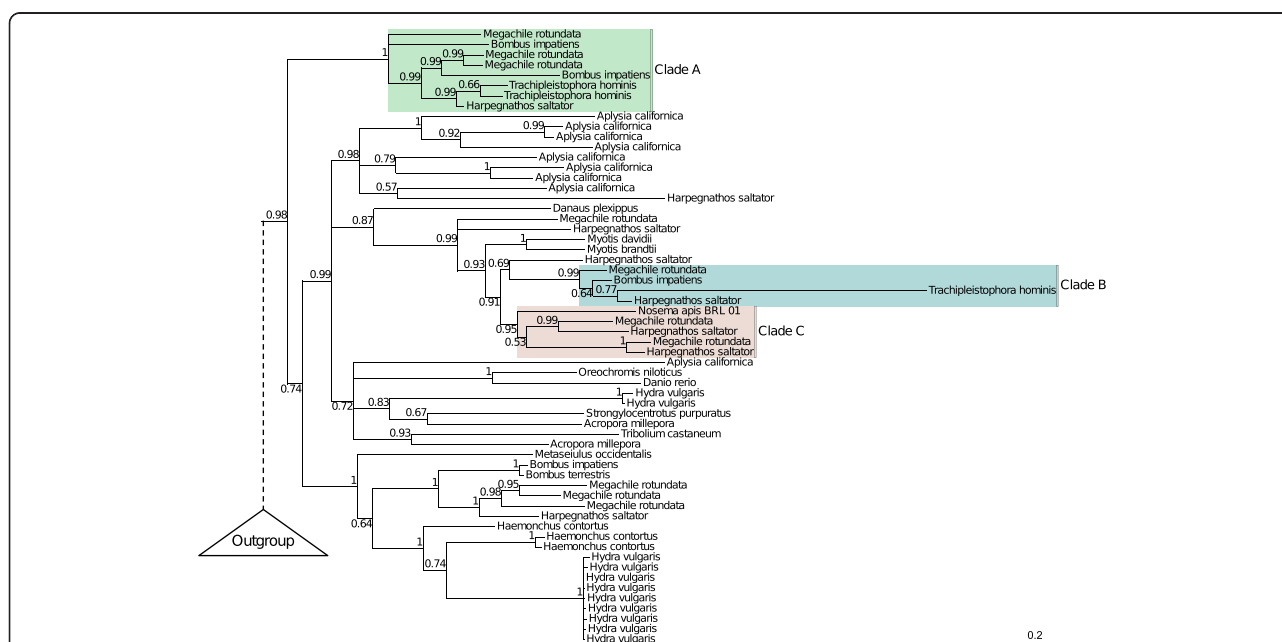


Fig. 7 Phylogenetic analysis of PiggyBac transposons suggests a natural insect host for *T. hominis*. A Bayesian phylogeny of *T. hominis* PiggyBac transposases inferred under the C20 model [117] in PhyloBayes [118]. Support values are Bayesian posterior probabilities, and branch lengths are proportional to the expected number of substitutions per site. The tree topology supports two recent, independent transfers of PiggyBac elements from hymenopteran insects into *T. hominis*. In both cases, the ant *Harpegnathos saltator* is recovered as the closest relative of the *T. hominis* sequence, although with variable posterior support. We also identify a transfer from insects into the microsporidian *Nosema apis*, a honeybee parasite. Clade A – An insect clade including two *T. hominis* PiggyBac elements, as identified in [10]. Clade B – A distinct clade of insect elements, including the newly discovered *T. hominis* PiggyBac element. Clade C – Clade including *Nosema apis* and insect PiggyBac elements

also provided some evidence for sex and recombination [81]. Although proliferating *T. hominis* cells can contain multiple nuclei, nuclear fusion has never been observed [6, 7]. Our RNA-Seq data presents an opportunity to investigate the ploidy of *T. hominis* and to test whether sexual reproduction may be possible.

We identified 7596 variant sites (polymorphisms) in the *T. hominis* transcriptome, with a total of 7654 possible variants. These included 7120 single nucleotide polymorphisms (SNPs), 314 insertions and 220 deletions. Plotting the allele frequency spectrum of the variations reveals a clear peak at a frequency of 0.5 (50 % reference genome allele, 50 % alternative allele) (Fig. 8a). Our *T. hominis* populations are likely to be clonal, both because their obligate intracellular lifecycle results in a population bottleneck in each generation, and also because our experimental isolate has been passaged repeatedly in cell culture. In addition, population-level variation would not be expected to give rise to a peak at 0.5, unless two distinct populations had somehow been maintained in a 50:50 ratio. Given these considerations, the simplest interpretation of the observed allele frequency spectrum is that the genome of *T. hominis* is diploid in at least some stage of its lifecycle and that, at least in the majority of cases, both alleles are expressed. *T. hominis* is unikaryotic – that is, it has one nucleus per spore – and so our results are consistent with analyses suggesting that other unikaryotic microsporidians are also diploid [16, 21, 22]. The diploidy of *T. hominis* and other unikaryotic Microsporidia supports the notion that diplokaryotic Microsporidia are likely to be tetraploid [83], containing two diploid nuclei as observed in the diplomonad *Giardia lamblia* [84]. The diploidy of *T. hominis* raises the possibility that it occasionally has sex, although the density of SNPs in our dataset was not sufficiently high to evaluate

the possibility of recombination or linkage disequilibrium. Although the *T. hominis* spore is unikaryotic, the intracellular stages of its lifecycle divide by a combination of binary division and plasmotomy, the division of a single cell producing multinucleate progeny [6, 7]. This raises the possibility that meiosis could still be triggered in multinuclear intracellular stages of the parasite lifecycle.

A total of 5496 SNPs were identified within annotated *T. hominis* ORFs, with 2175 non-synonymous and 2933 synonymous changes. SNPs were identified in 1551 ORFs in total, with 1071 of these ORFs including non-synonymous changes, leading to possible protein variants. Although we cannot exclude the possibility that some non-synonymous mutations might be beneficial, we expect the majority to be deleterious, particularly given the frequent genetic bottlenecks experienced by our artificially maintained experimental population. 388 changes were more severe in nature, either including frameshifts or alterations to start and stop codons. In principle, the deleterious effects of these mutations could be suppressed in a diploid organism by preferential expression of the reference allele. However, the peak at 0.5 is still observed in the frequency spectrum for non-synonymous alleles (Fig. 8b), implying that synonymous, non-synonymous and reference alleles are all expressed at similar levels. 34 % of all SNPs occurred in “core” microsporidian genes – those shared among at least 9 of 11 sequenced Microsporidia and predicted to play an important role in the parasite. The expression of these variants is perhaps surprising due to their potential impact on protein function. However, it is important to remember that this study examines a single population of *T. hominis*, and that many of these SNPs may be en route to elimination by negative selection. Another possibility is that the high levels of expression we observed

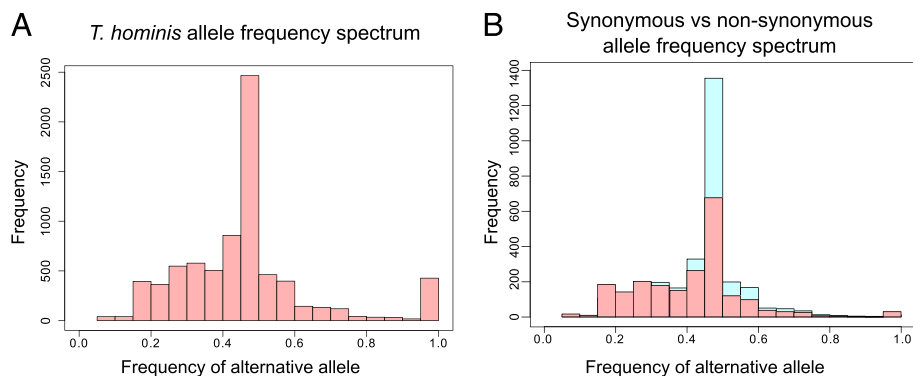


Fig. 8 The allele frequency spectrum of SNPs in expressed *T. hominis* transcripts suggests a diploid phase to the parasite lifecycle. **a** The allele frequency spectrum (reference and alternative alleles) for *T. hominis* SNPs. The clear peak at an allele frequency of 0.5 suggests that *T. hominis* is diploid. **b** Allele frequency spectra with synonymous and non-synonymous SNPs plotted separately. The distributions are similar, providing no evidence for reduced expression of non-synonymous alleles

for key molecular chaperones may help to suppress the phenotypic effect of these deleterious mutations, as has previously been reported for intracellular bacteria [46].

Response of the RK-13 cell line to infection with *T. hominis*

Recent studies have begun to shed light on the mechanisms by which microsporidians exploit their hosts [13, 14, 85, 86], but we still know almost nothing about the host response to infection at the molecular level. We quantified host gene expression to identify genes and pathways that were differentially expressed in RK-13 cells during infection with *T. hominis* compared to uninfected cells (Additional file 9: Table S4). These analyses provide a first snapshot of the impact of *T. hominis* infection on the host cell transcriptome. We identified 1734 transcripts that showed significant changes in expression in the RK-13 cell line during infection with *T. hominis* (Additional file 10: Table S5). KEGG categories [87] were assigned to these transcripts using the KOBAS annotation pipeline [88]; nine KEGG categories were enriched for differentially expressed transcripts (Additional file 11: Table S6); that is, these categories contained more genes whose expression levels changed in response to infection than would be expected by chance, allowing us to explore the general effects of infection on the host cell.

Our analysis suggests that the host experiences a generalised cellular shutdown in response to *T. hominis* infection, with the down-regulation of the great majority of genes involved in the KEGG pathways for the cell cycle, meiosis, DNA replication, and ribosome biogenesis (Additional file 12: Figure S6). Several other host cell pathways contained a mixture of both up-regulated and down-regulated genes relative to the uninfected control; these pathways are clearly disrupted during *T. hominis* infection, but the overall effect on the host is difficult to predict from changes in transcript abundance. The pathways included those involved in focal adhesion, extracellular matrix-receptor interactions, renal cell carcinoma, oxidative phosphorylation, and pyrimidine biosynthesis. Although these patterns are complex and difficult to interpret, they were also consistent across our three biological replicates, and therefore represent reproducible perturbations of host cell pathways.

The only host cell pathways in which the majority of changes were up-regulations relative to the control were amino sugar and nucleotide sugar metabolism (Additional file 12: Figure S6), potentially leading to increased production of nucleotide sugars by the host. Similar changes to silkworm metabolism were observed during infection with *Nosema bombycis*, suggesting that this may be common feature of microsporidian infection [18]. Candidate nucleotide sugar importers have been

identified in *T. hominis* [10] and other Microsporidia, consistent with the idea that microsporidians might manipulate host metabolism to increase production of required substrates. One potential mechanism for the manipulation of host metabolism that has already been proposed is the secretion of hexokinase into the host cell [16]. As in mammalian cancer cells [89], Microsporidia infected RK-13 cells had a modulated hexokinase expression profile compared to healthy cells, possibly to the benefit of the intracellular parasite. The two hexokinase isozymes (HKI and HKII) with the highest affinity for glucose in mammals [90] were significantly differentially expressed in the host during infection, with an increase in HKI and a decrease in HKII. HKI is believed to have a primarily catabolic function, driving glycolysis and ATP production [90] – an essential molecule for *T. hominis* growth and replication. These changes in gene expression draw striking parallels to other host-parasite systems, where complex changes in host energy metabolism [91, 92] and pyrimidine biosynthesis [93] are associated with infection.

A number of pathway regulators were also differentially expressed in RK-13 cells during *T. hominis* infection, providing insights into the potential mechanisms that might underpin some of the observed changes in transcript levels of metabolic genes. We observed significant up-regulation of host 5'-AMP-activated protein kinase catalytic subunit alpha-2 (PRKAA2) and peroxisome proliferator-activated receptor gamma co-activator 1-alpha (PPARGC1A) during *T. hominis* infection, both of which are reported to promote energy metabolism and mitochondrial biogenesis [94, 95]. PRKAA2 is additionally implicated in shutting down ATP-consuming pathways including cell proliferation [94], consistent with our observation of decreased expression of genes in this pathway.

Linked to the above inference of increased host ATP production coupled with reduced consumption, we also observed a significant decrease in the expression of host pyruvate dehydrogenase lipoamide kinase isozyme 4 (PDK4) during infection. This kinase represses metabolism through the phosphorylation and inactivation of pyruvate dehydrogenase, the enzyme that converts pyruvate to acetyl-coA, thereby linking glycolysis and the citric acid cycle [96]. Decreased expression of PDK4 would lead to increased activity of pyruvate dehydrogenase, promoting citric acid cycle-based metabolism and, under normal oxygen conditions, increased ATP production [96, 97]. The increased glucose required to support the elevated metabolic demand might be acquired by increased import since the glucose transporter GLUT9 [98] is highly up-regulated transporter during infection by *T. hominis*.

Changes in expression in the ubiquitination pathway have been recently implicated in the immune response

of *C. elegans* to *N. parisii* infection [19], raising the possibility that the ubiquitin system may be part of a common host response to microsporidian infection. Although the expression levels of several host ubiquitination genes were altered upon *T. hominis* infection, the pathway as a whole was not enriched in differentially expressed genes in our analysis, suggesting that this does not represent a major component of the host cell response to infection by *T. hominis*.

Conclusions

Our transcriptomics data was highly reproducible for parasite and host, and confirmed that *T. hominis*, with ~3150 genes, has one of the largest coding capacities among microsporidians [10, 20]. Although our data rejected some of the shortest predicted gene models, this was compensated by the identification of genes that were previously missed by the genome annotation. Some of these, including transporters that may acquire pyrimidines and enzymes that function in chitin biosynthesis, may plug what were previously considered to be gaps in the metabolic capacity of the parasite.

Gene expression for the parasite was highly biased towards growth and replication, consistent with published microscopic data [6, 7] demonstrating rapid parasite proliferation after infection. Intriguingly, a proportion of the highly expressed transcripts are encoded by conserved microsporidian genes of unknown function, suggesting there is much still to discover about the core biology of these highly successful parasites. Expression within expanded gene families, including key transport proteins, was highly variable. In most cases the most highly conserved members of gene families were also the most highly expressed, consistent with evolutionary models in which duplication can free individual paralogues to diversify in function while preserving the ancestral function in the conserved copy. The expression of genes confined to *T. hominis* and its close relative *Vavraia culicis* was also more heterogeneous than was observed for core genes. Some of this lineage-specific innovation was highly expressed – in particular a membrane protein of unknown function – but much of it was not. This class of genes is also enriched for signal peptides [11] suggesting that some may be secreted or exposed on the surface of the parasite where they can interact with host targets. Our results contribute to a growing body of work supporting the idea that the evolution of contemporary microsporidian genomes is highly dynamic and innovative, and that while the initial transition to intracellular parasitism catalysed a drastic reduction in genome size and coding capacity shared by all microsporidians, important lineage-specific differences continue to evolve.

Our data strongly suggest that *T. hominis* is diploid and demonstrate the presence of a large number of non-synonymous SNPs, many of which are expected to be deleterious, that are equally distributed between alleles. Many of these SNPs may eventually be eliminated by negative selection but, as already suggested for intracellular bacteria [46–49], the high levels of chaperonin expression that we observed may also suppress the phenotypic effects of these deleterious mutations in Microsporidia [41]. It has been demonstrated that artificially infected mosquitoes can support the replication of *T. hominis* [74], but the natural host of this opportunistic pathogen of humans is currently unknown. We identified transcripts from a novel PiggyBac element that, together with a previously identified element of insect origin [10], strongly suggest that the natural host for *T. hominis* belongs to the hymenoptera.

The response of eukaryotic host cells to microsporidian infection is only just beginning to be investigated. Our data, which were highly reproducible between biological and technical replicates, suggest a generalized cellular shutdown by infected cells compared to uninfected rabbit kidney cells. Several other host pathways displayed a reproducible mixture of up-regulated and down-regulated genes relative to the uninfected control; these pathways are clearly disrupted but the overall effect on the host and its relationship to the activities of the parasite are difficult to predict based solely upon these data. We did observe an up-regulation of host amino and nucleotide sugar metabolism: this has also been reported for silkworms infected with *Nosema bombycis*. These are among substrates predicted to be imported by microsporidians to plug gaps in their reduced metabolism [11, 16], so it is possible that these changes are to the benefit of the parasites. There is some evidence that host ATP production might be increased in combination with reduced host energy consumption. This could potentially benefit a parasite that is dependent on the host cell for most of its ATP and purine nucleotides for DNA and RNA biosynthesis [14].

Methods

Culture of *T. hominis* infected RK-13 cells

T. hominis was grown in co-culture with RK-13 cells [6] grown in minimal essential medium (MEM) containing 10 % heat inactivated foetal calf serum and antibiotics (Penicillin/Streptomycin (0.1 mg/ml), Ampicillin B (1 µg/ml) and Kanamycin sulphate (0.1 mg/ml)). A single 175 cm² flask of RK-13 cells was grown to confluence (defined as a continuous cell monolayer) and split into three separate flasks. These flasks were raised in parallel until confluence, when each was again split into two centrifuge tubes. Samples were spun at 400 g and trypsin removed. The cells were resuspended in 5 ml MEM.

Spores were freshly harvested from 20 flasks of *T. hominis*-infected RK-13 cells and resuspended in 400 µl PBS ($\sim 2.3 \times 10^7$ spores/ml). 100 µl of this spore suspension was added to one of the two centrifuge tubes containing RK-13 cells. Cells were incubated for 10 min before seeding to a new 175 cm² flask containing 40 ml MEM.

Flasks of uninfected and *T. hominis* infected RK-13 cells were raised in parallel for 7 days post inoculation. During this time they were trypsinised and split twice in order to boost levels of infection. Two days after the final trypsinisation the cells were harvested in RNeasy lysis reagent and immediately frozen at -20°C . At this stage, approximately 60 % of cells in flasks to which spores had been added exhibited signs of infection.

Preparation of RNA for sequencing

Cells suspended in RNeasy lysis reagent (Qiagen) were thawed and pelleted by centrifugation at 400 g. Total RNA was purified from each sample using the standard TRIzol (Invitrogen) extraction protocol, with the addition of a bead beating step (three times for 20 s at 5 m/s, with 10 mins incubation on ice between beating). An additional cleanup step using the GeneJet RNA purification kit (Thermo Scientific) was added to remove residual organic solvents from the purified total RNA. The RNA integrity and concentration was assessed using the Agilent RNA 6000 Nano Kit on the Agilent 2100 BioAnalyser. PolyA RNA was isolated from 5 µg of purified total RNA. Libraries were prepared using the ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Epicentre Biotechnologies) and sequenced on the Illumina HiSeq 2500 in Rapid-Run mode, producing non-strand-specific 100 bp single-ended reads, with each library sequenced on two different lanes of the sequencer.

Processing and analysis of RNA-seq data

The *Trachipleistophora hominis* genome and annotation [10] were obtained from NCBI, whilst the genome and annotation of the European rabbit (*Oryctolagus cuniculus*: GCA_000003625.1) were obtained from the Ensembl database. Bowtie2 [99] was used to separately index the genomes of *T. hominis* and *O. cuniculus*. Quality control on the raw RNA sequencing reads was performed using FastQC [100] and Illumina sequencing adapters and low quality bases were trimmed using fastq-mcf [101]. In order to quantify the expression levels of *T. hominis* transcripts, and for novel transcript discovery, TopHat2 [102] was used to map quality-filtered reads from each infected sample to the *T. hominis* genome. Transcripts were assembled using cufflinks [103]. The final transcriptome assembly was generated using cuffmerge [103]. All sequence data associated with this project has been deposited at NCBI under the BioProject ID PRJNA 278775. Linear mixed-effect models were used to assess

differential expression between different categories of genes within the parasite transcriptome. We fit functional category as a fixed effect, with random effects for gene, technical replicate, and biological replicate, and used log FPKM values as the response variable.

Transcripts that mapped to unannotated regions of the *T. hominis* genome were screened for potential ORFs by using BLASTx to search against the nr database with an E-value threshold of 0.01. SNPs were identified using SamTools mpileup and bcftools [104]. Vcfutils.pl varFilter was applied under default settings to remove low quality SNPs, with the addition of a minimum read depth of 10 [105]. An additional filter was applied to remove bases with low mapping quality scores. Values of 20, 40, 60 and 80 were tested. In all cases application of the filter reduced non-peak (0.5 frequency) signals while retaining the overall distribution of the allele frequency spectrum. We used a 60 as a balance between stringently filtering out low quality mapping and retaining data. The location of SNPs relative to annotated *T. hominis* genes and the impact of SNPs on protein-coding sequences were assessed using SNPeff [106] and processed using SNPsift [107].

To maximise the data available for intron detection, reads from all samples including *T. hominis*-infected cells were pooled and mapped onto the *T. hominis* reference genome using TopHat2 [102]. The intron junctions identified from this mapping were manually filtered so as to retain only those junctions covered by more than one read. Overlapping junctions were merged.

For quantification of *O. cuniculus* transcripts, TopHat2 [102] was used to map reads from all samples from one lane of the sequencer to the *O. cuniculus* genome, and transcripts were assembled and quantified using cufflinks. The abundance of transcripts in the three flasks of uninfected RK-13 cells was compared to that in the three flasks of *T. hominis* infected RK-13 cells using cuffdiff [108]. All RNA sequencing results were analysed in R using the cummeRbund package [109]. KEGG pathway [87] assignment for significantly differentially expressed genes and pathway enrichment analysis was carried out using KOBAS 2.0 [88].

Phylogenetic analysis of PiggyBac elements

We BLASTed *T. hominis* PiggyBac elements THOM_1159, THOM_1429 and the additional family member newly identified in our transcriptome against the nr database, retrieving the top 100 significant hits with an E-value of less than 0.05. Duplicate hits were manually removed before sequences were aligned using M-Coffee [110], combining the results of alignments using Muscle [111], Mafft [112], ProbCons [113], PCMA [114], and Fsa [115]. Poorly-aligning regions were identified and removed using trimAl [116]. Our phylogeny was inferred

using the C20 model [117] implemented in PhyloBayes 3.3 [118].

Assessing heterogeneous expression in duplicated gene families

The expanded gene families in *T. hominis* and other Microsporidia identified by Nakjang et al. [11] where *T. hominis* had at least two duplicate copies were investigated to compare the expression of individual genes. To identify the *T. hominis* gene families showing the greatest level of between-paralogue expression level divergence, we normalised the expression level of each gene by the average level of expression for the family to which it belonged, then took the standard deviation of these values for each family. We then ranked families by this score to identify the families with the greatest within-family divergence.

Microscopy of *T. hominis* infected RK-13 cells

For phase contrast microscopy cells were fixed in a 50 % methanol 50 % acetone solution. After washing in water slides were mounted in Mowiol containing p-Phenylenediamine (0.01 %) which was allowed to set overnight at room temperature. and mounted in the same way. Microscopy was performed using the Zeiss Axio Imager II (Upright) in structured illumination (apotome) mode at the Newcastle University Bio-Imaging unit.

Ethical approval

T. hominis was maintained in the RK-13 cell line so no ethical approval was required for this study.

Availability of supporting data

All sequence data associated with this project has been deposited at NCBI under the BioProject ID PRJNA 278775, and the *Trachipleistophora hominis* genome assembly (GCA_000316135.1) has been updated with the new gene models identified here.

Additional files

Additional file 1: Figure S1. Robustness and reproducibility of transcriptomic analysis for RK-13 cells during intracellular infection. Pairwise comparison of log₁₀ FPKM values in rabbit kidney cell transcript quantification, as outlined in Fig. 2a for parasite transcripts. In this case biological replicates of uninfected RK-13 cells (samples labelled RK) are compared to those for *T. hominis* infected RK-13 cells (samples labelled TH). (TIFF 6369 kb)

Additional file 2: Table S1. Expression levels of *T. hominis* transcripts. These include novel transcripts identified in this study (Locus tag labelled as "NA"). Genes annotated in the *T. hominis* genome but not detected during this study are also included for reference. In some cases multiple locus tags map to a single transcript, as discussed in the main text. (XLSX 486 kb)

Additional file 3: Table S2. Predicted *T. hominis* intron junctions. The genomic location and sequencing depth of all intron junctions identified in the *T. hominis* genome by TopHat (see Methods). (XLSX 11 kb)

Additional file 4: Figure S2. Alignment of a selection of identified intron sequences. An alignment of novel *T. hominis* intron sequences predicted from our transcriptome, suggesting no conserved intron motif. The alignment was generated using MUSCLE [111] and visualised using SeaView [120]. (PDF 62 kb)

Additional file 5: Figure S3. Overall expression profile of the *T. hominis* transcriptome. Distribution of ranked FPKM values for the *T. hominis* transcriptome. The 95th percentile (2240 FPKM) is marked by a blue line and the mean expression value (705 FPKM) is marked by a red line. (PDF 30 kb)

Additional file 6: Table S3. Heterogeneity in expression of duplicated gene families. All duplicated gene families [11] containing at least two paralogues in *T. hominis* ranked according to their expression heterogeneity index (see Methods). (XLSX 25 kb)

Additional file 7: Figure S4. Heterogeneous expression in expanded microsporidian gene families. Distribution of ranked heterogeneity index (see Methods). The red line denotes the inflection point above which we considered gene families to show high levels of expression heterogeneity. (PDF 18 kb)

Additional file 8: Figure S5. Phylogenetic analysis of PiggyBac transposons suggests a natural insect host for *T. hominis*—full tree. An expanded version of the tree from Figure 4 including GI accessions for all included sequences. The accession number of the newly identified *T. hominis* sequence (Clade B) is XLOC_002128. (PDF 22 kb)

Additional file 9: Table S4. Expression levels of RK-13 transcripts. Where possible these were annotated with Ensembl gene names. For unannotated transcripts identified in this study Ensembl Gene IDs are labelled as "NA". Genes annotated in the *O. cuniculus* genome with no detectable expression during this study are also included for reference. Due to overlapping genes and alternative splicing, some transcriptome gene IDs are associated with multiple Ensembl gene IDs and/or multiple Ensembl Protein IDs. (XLSX 8052 kb)

Additional file 10: Table S5. RK-13 transcripts identified as significantly differentially expressed during infection. Genes identified by cuffdiff [108] as significantly differentially expressed in the RK-13 host cell during *T. hominis* infection. Q-values are P-values that have been corrected for multiple testing using the false discovery rate method [108]. (XLSX 345 kb)

Additional file 11: Table S6. Testing for enrichment of differentially expressed genes in KEGG pathways. Fischer's exact test with Benjamini and Hochberg correction [121] for the enrichment of genes identified as significantly differentially expressed in KEGG pathways in the RK-13 cell line. (XLSX 70 kb)

Additional file 12: Figure S6. Host cell gene expression changes in KEGG pathways enriched for differentially expressed genes. Genes highlighted in cyan are significantly up-regulated upon infection with *T. hominis*, genes in red are down-regulated, and the expression of genes in purple does not significantly change. Genes were assigned to KEGG pathways [87] using the KOBAS annotation pipeline [88], and colours representing differential expression were assigned using the KEGG web server [122]. (PDF 613 kb)

Abbreviations

FPKM: Fragments per kilobase of transcript per million mapped reads; NTT: Nucleotide transport protein; ORF: Open reading frame; RK: Rabbit kidney; SNP: Single nucleotide polymorphism; UTR: Untranslated region.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AKW, TAW, RPH, and TME designed the experiments. AKW carried out the experiments and analysed the data. BAPW and KAM provided help and guidance with sample preparation and the transcriptomic analyses. All

authors contributed to the interpretation of the results. AKW and TAW wrote the paper, with editing from RPH and TME. All authors have read and approved the manuscript.

Acknowledgements

This work was supported by a BBSRC studentship to A.K.W., a Marie Curie Intra-European postdoctoral fellowship to T.A.W., and the European Research Council Advanced Investigator Programme (ERC- 2010-AdG-268701) and the Wellcome Trust (grant number 045404) to T.M.E. K.A.M. was supported by a Wellcome Trust Institutional Strategic Support Award (WT097835MF). We thank M. E. Geggie and E. Kozhevnikova for assistance with cell culture and lab reagents, Dr. S. E. Campbell for assistance in RNA purification and sample preparation, Dr. K. Moore and A. Farbos at the Exeter Sequencing Service for assistance in Illumina sequencing and library preparation, Dr T. A. Booth for microscopy training and K. Sendra, A. Farbos, and Dr. S. E. Heaps for helpful discussions. We also thank the two anonymous reviewers for their insightful comments.

Author details

¹Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. ²Biosciences, College of Life and Environmental Sciences, University of Exeter, Devon, UK.

Received: 7 April 2015 Accepted: 6 October 2015

Published online: 21 November 2015

Bibliography

- Hirt RP, Logsdon JM, Healy B, Dorey MW, Doolittle WF, Embley TM. Microsporidia are related to Fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. *Proc Natl Acad Sci*. 1999;96:580–5.
- James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, et al. Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. *Curr Biol*. 2013;23:1548–53.
- Vávra J, Lukeš J. Microsporidia and "the art of living together". *Adv Parasitol*. 2013;82:253–319.
- Vidau C, Diogon M, Aufaure J, Fontbonne R, Viguès B, Brunet J-L, et al. Exposure to sublethal doses of fipronil and thiacloprid highly increases mortality of honeybees previously infected by *Nosema ceranae*. *PLoS One*. 2011;6:e21550.
- Didier ES, Weiss LM. Microsporidiosis: not just in AIDS patients. *Curr Opin Infect Dis*. 2011;24:490–5.
- Hollister WS, Canning EU, Weidner E, Field AS, Kench J, Marriott DJ. Development and ultrastructure of *Trachipleistophora hominis* n.g., n.sp. after in vitro isolation from an AIDS patient and inoculation into athymic mice. *Parasitology*. 1996;112(Pt 1):143–54.
- Field AS, Marriott DJ, Milliken ST, Brew BJ, Canning EU, Kench JG, et al. Myositis associated with a newly described microsporidian, *Trachipleistophora hominis*, in a patient with AIDS. *J Clin Microbiol*. 1996;34:2803–11.
- Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*. 2001;414:450–3.
- Corradi N, Pombert J-F, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat Commun*. 2010;1:77.
- Heinz E, Williams TA, Nakjang S, Noël CJ, Swan DC, Goldberg AV, et al. The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. *PLoS Pathog*. 2012;8:e1002979.
- Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, et al. Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. *Genome Biol Evol*. 2013;5:2285–303.
- Haag KL, James TY, Pombert J-F, Larsson R, Schaer TMM, Refardt D, et al. Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *Proc Natl Acad Sci*. 2014;111(43):15480–5.
- Tsao AD, Kunji ERS, Goldberg AV, Lucocq JM, Hirt RP, Embley TM. A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. *Nature*. 2008;453:553–6.
- Heinz E, Hacker C, Dean P, Mifsud J, Goldberg AV, Williams TA, et al. Plasma Membrane-Located Purine Nucleotide Transport Proteins Are Key Components for Host Exploitation by Microsporidian Intracellular Parasites. *PLoS Pathog*. 2014;10:e1004547.
- Grisdale CJ, Bowers LC, Didier ES, Fast NM. Transcriptome analysis of the parasite *Encephalitozoon cuniculi*: an in-depth examination of pre-mRNA splicing in a reduced eukaryote. *BMC Genomics*. 2013;14:207.
- Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, et al. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res*. 2012;22:2478–88.
- Campbell SE, Williams TA, Yousuf A, Soanes DM, Paszkiewicz KH, Williams BAP. The genome of *Spraguea lophii* and the basis of host-microsporidian interactions. *PLoS Genet*. 2013;9:e1003676.
- Ma Z, Li C, Pan G, Li Z, Han B, Xu J, et al. Genome-wide transcriptional response of silkworm (*Bombyx mori*) to infection by the microsporidian *Nosema bombycis*. *PLoS One*. 2013;8:e84137.
- Bakowski MA, Desjardins CA, Smelkinson MG, Dunbar TA, Lopez-Moyado IF, Rifkin SA, Cuomo CA, Troemel ER. Ubiquitin-Mediated Response to Microsporidia and Virus Infection in *C. elegans*. *PLoS Pathog*. 2014;10:e1004200.
- Peyretillade E, Boucher D, Parisot N, Gasc C, Butler R, Pombert J-F, Lerat E, Peyret P. Exploiting the architecture and the features of the microsporidian genomes to investigate diversity and impact of these parasites on ecosystems. *Heredity (Edinb)*. 2014, ePub.
- Haag KL, Trautnecker E, Ebert D. Single-nucleotide polymorphisms of two closely related microsporidian parasites suggest a clonal population expansion after the last glaciation. *Mol Ecol*. 2013;22:314–26.
- Selman M, Sak B, Kváč M, Farinelli L, Weiss LM, Corradi N. Extremely reduced levels of heterozygosity in the vertebrate pathogen *Encephalitozoon cuniculi*. *Eukaryot Cell*. 2013;12:496–502.
- Vávra J, Becnel JJ. *Vavraia culicis* (Weiser, 1947) Weiser, 1977 revisited: cytological characterisation of a *Vavraia culicis*-like microsporidium isolated from mosquitoes in Florida and the establishment of *Vavraia culicis floridensis* subsp. n. *Folia Parasitol (Praha)*. 2007;54:259–71.
- Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, et al. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat Commun*. 2015;6: 7121. doi: 10.1038/ncomms8121.
- Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. *Bioinformatics*. 2009;25:1335–7.
- Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, et al. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res*. 2013;41(Database issue):D226–32.
- Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 2005;21:951–60.
- Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33(Web Server issue):W244–8.
- Muralidharan V, Goldberg DE. Asparagine repeats in *Plasmodium falciparum* proteins: good for nothing? *PLoS Pathog*. 2013;9:e1003488.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8:785–6.
- Park MH, Wolff EC, Folk JE. Hypusine: its post-translational formation in eukaryotic initiation factor 5A and its potential role in cellular regulation. *Biofactors*. 1993;4:95–104.
- Bassler J, Kallas M, Pertschy B, Ulbrich C, Thoms M, Hurt E. The AAA-ATPase Rea1 drives removal of biogenesis factors during multiple stages of 60S ribosome assembly. *Mol Cell*. 2010;38:712–21.
- DeLabre ML, Kessl J, Karamanou S, Trumpower BL. RPL29 codes for a non-essential protein of the 60S ribosomal subunit in *Saccharomyces cerevisiae* and exhibits synthetic lethality with mutations in genes for proteins required for subunit coupling. *Biochim Biophys Acta*. 2002;1574:255–61.
- Weiss LM, Delbac F, Hayman JR, Pan G, Dang X, Zhou Z. The Microsporidian Polar Tube and Spore Wall. In: Weiss LM, Becnel JJ, editors. *Microsporidia Pathog Oppor*. 1st ed. Iowa: John Wiley & Sons, Inc; 2014. p. 1–70.
- Zachara NE, Hart GW. O-GlcNAc as a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim Biophys Acta*. 2004;1673:13–28.
- Slawson C, Housley MP, Hart GW. O-GlcNAc cycling: how a single sugar post-translational modification is changing the way we think about signaling networks. *J Cell Biochem*. 2006;97:71–83.
- Stokes MJ, Güther MLS, Turnock DC, Prescott AR, Martin KL, Alpey MS, et al. The synthesis of UDP-N-acetylglucosamine is essential for bloodstream form *Trypanosoma brucei* in vitro and in vivo and UDP-N-acetylglucosamine

- starvation reveals a hierarchy in parasite protein glycosylation. *J Biol Chem*. 2008;283:16147–61.
38. Urbaniaik MD, Collie IT, Fang W, Aristotelous T, Eskilsson S, Raimi OG, et al. A novel allosteric inhibitor of the uridine diphosphate N-acetylglucosamine pyrophosphorylase from *Trypanosoma brucei*. *ACS Chem Biol*. 2013;8:1981–7.
 39. Williams BAP, Slamovits CH, Patron NJ, Fast NM, Keeling PJ. A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc Natl Acad Sci U S A*. 2005;102:10936–41.
 40. Corradi N, Gangaeva A, Keeling PJ. Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospira locustae* and *Encephalitozoon cuniculi*. *Genomics*. 2008;91:388–93.
 41. Gill EE, Becnel JJ, Fast NM. ESTs from the microsporidian *Edhazardia aedis*. *BMC Genomics*. 2008;9:296.
 42. Zorio DA, Cheng NN, Blumenthal T, Spieth J. Operons as a common form of chromosomal organization in *C. elegans*. *Nature*. 1994;372:270–2.
 43. Lee RCH, Gill EE, Roy SW, Fast NM. Constrained intron structures in a microsporidian. *Mol Biol Evol*. 2010;27:1979–82.
 44. Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R, Bähler J. Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells. *Cell*. 2012;151:671–83.
 45. Henderson B, Allan E, Coates ARM. Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. *Infect Immun*. 2006;74:3693–706.
 46. Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E. Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature*. 2002;417:398.
 47. Williams TA, Fares MA. The effect of chaperonin buffering on protein evolution. *Genome Biol Evol*. 2010;2:609–19.
 48. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 2012;10:13–26.
 49. Bogumil D, Dagan T. Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry*. 2012;51:9941–53.
 50. Liberek K, Lewandowska A, Zietkiewicz S. Chaperones in control of protein disaggregation. *EMBO J*. 2008;27:328–35.
 51. Dolgikh VV, Sokolova JJ, Issi IV. Activities of enzymes of carbohydrate and energy metabolism of the spores of the microsporidian, *Nosema grylli*. *J Eukaryot Microbiol*. 1997;44:246–9.
 52. Tsunehiro F, Junichi N, Narimichi K, Kazutada W. Isolation, overexpression and disruption of a *Saccharomyces cerevisiae* YNK gene encoding nucleoside diphosphate kinase. *Gene*. 1993;129:141–6.
 53. Vértessy BG, Tóth J. Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Acc Chem Res*. 2009;42:97–106.
 54. Gowri VS, Ghosh I, Sharma A, Madhubala R. Unusual domain architecture of aminoacyl tRNA synthetases and their paralogs from *Leishmania major*. *BMC Genomics*. 2012;13:621. doi: 10.1186/1471-2164-13-621.
 55. Manhas R, Tripathi P, Khan S, Sethu Lakshmi B, Lal SK, Gowri VS, et al. Identification and functional characterization of a novel bacterial type asparagine synthetase A: a tRNA synthetase paralog from *Leishmania donovani*. *J Biol Chem*. 2014;289:12096–108.
 56. Loureiro I, Faria J, Clayton C, Ribeiro SM, Roy N, Santarém N, et al. Knockdown of Asparagine Synthetase A Renders *Trypanosoma brucei* Auxotrophic for Asparagine. *PLoS Negl Trop Dis*. 2013;7:e2578.
 57. Lu SC. Regulation of glutathione synthesis. *Mol Aspects Med*. 2009;30:42–59.
 58. Biron DG, Agnew P, Marché L, Renault L, Sidobre C, Michalakakis Y. Proteome of *Aedes aegypti* larvae in response to infection by the intracellular parasite *Vavraia culicis*. *Int J Parasitol*. 2005;35:1385–97.
 59. Duncan AB, Agnew P, Noel V, Demetree E, Seveno M, Brizard J-P, et al. Proteome of *Aedes aegypti* in response to infection and coinfection with microsporidian parasites. *Ecol Evol*. 2012;2:681–94.
 60. Panek J, El Alaoui H, Mone A, Urbach S, Demetree E, Texier C, et al. Hijacking of Host Cellular Functions by an Intracellular Parasite, the Microsporidian *Anncalia algerae*. *PLoS One*. 2014;9:e100791.
 61. Vanaporn M, Wand M, Michell SL, Sarkar-Tyson M, Ireland P, Goldman S, et al. Superoxide dismutase C is required for intracellular survival and virulence of *Burkholderia pseudomallei*. *Microbiology*. 2011;157(Pt 8):2392–400.
 62. Goldberg AV, Molik S, Tsaousis AD, Neumann K, Kuhnke G, Delbac F, et al. Localization and functionality of microsporidian iron-sulphur cluster assembly proteins. *Nature*. 2008;452:624–8.
 63. Zhang Y, Lyver ER, Nakamaru-Ogiso E, Yoon H, Amutha B, Lee D-W, et al. Dre2, a Conserved Eukaryotic Fe/S Cluster Protein, Functions in Cytosolic Fe/S Protein Biogenesis. *Mol Cell Biol*. 2008;28:5569–82.
 64. Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
 65. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39(Web Server issue):W29–37.
 66. Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B, et al. Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *PLoS Pathog*. 2009;5:e1000466.
 67. Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, et al. Proto-genes and de novo gene birth. *Nature*. 2012;487:370–4.
 68. Becnel JJ, Andreadis TJ. Microsporidia in Insects. In: Weiss LM, Becnel JJ, editors. *Microsporidia Pathog Oppor*. 2014.
 69. Ohno S. Evolution by Gene Duplication. New York: Springer; 1970.
 70. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 1999;151:1531–45.
 71. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*. 2004;428:617–24.
 72. Conant GC, Wagner A. Asymmetric sequence divergence of duplicate genes. *Genome Res*. 2003;13:2052–8.
 73. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet*. 2008;9:938–50.
 74. Weidner E, Canning EU, Rutledge CR, Meek CL. Mosquito (Diptera: Culicidae) Host Compatibility and Vector Competency for the Human Myositic Parasite *Trachipleistophora hominis* (Phylum Microspora). *J Med Entomol*. 1999;36:522–5.
 75. Cary LC, Goebel M, Corsaro BG, Wang H-G, Rosen E, Fraser MJ. Transposon mutagenesis of baculoviruses: Analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*. 1989;172:156–69.
 76. Mitra R, Li X, Kapusta A, Mayhew D, Mitra RD, Feschotte C, et al. Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc Natl Acad Sci U S A*. 2013;110:234–9.
 77. Zander E. Tierische Parasiten als Krankheitsserreger bei der Biene. *Münchener Bienenzeitung*. 1909;31:196–204.
 78. Pan G, Xu J, Li T, Xia Q, Liu S-L, Zhang G, et al. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC Genomics*. 2013;14:186.
 79. Guo X, Gao J, Li F, Wang J. Evidence of horizontal transfer of non-autonomous Lep1 Helitrons facilitated by host-parasite interactions. *Sci Rep*. 2014;4:5119.
 80. Parisot N, Pelin A, Gasc C, Polonais V, Belkorchia A, Panek J, et al. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. *Genome Biol Evol*. 2014;6:2289–300.
 81. Ironside JE. Multiple losses of sex within a single genus of Microsporidia. *BMC Evol Biol*. 2007;7:48.
 82. Hazard EI, Brookbank JW. Karyogamy and meiosis in an *Amblyospora* sp. (Microspora) in the mosquito *Culex salinarius*. *J Invertebr Pathol*. 1984;44:3–11.
 83. Pelin A, Selman M, Aris-Brosou S, Farinelli L, Corradi N. Genome analyses suggest the presence of polyploidy and recent human-driven expansions in eight global populations of the honeybee pathogen *Nosema ceranae*. *Environ Microbiol*. 2015. doi:10.1111/1462-2920.12883.
 84. Bernander R, Palm JED, Svard SG. Genome ploidy in different stages of the *Giardia lamblia* life cycle. *Cell Microbiol*. 2001;3:55–62.
 85. Troemel ER, Félix M-A, Whiteman NK, Barrière A, Ausubel FM. Microsporidia are natural intracellular parasites of the nematode *Caenorhabditis elegans*. *PLoS Biol*. 2008;6:2736–52.
 86. Estes KA, Szumowski SC, Troemel ER. Non-Lytic. Actin-Based Exit of Intracellular Parasites from *C. elegans* Intestinal Cells *PLoS Pathog*. 2011;7:e1002227.
 87. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
 88. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res*. 2011;39(Web Server issue):W316–22.
 89. Mathupala SP, Ko YH, Pedersen PL. The pivotal roles of mitochondria in cancer: Warburg and beyond and encouraging prospects for effective therapies. *Biochim Biophys Acta*. 2010;1797:1225–30.
 90. Wilson JE. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *J Exp Biol*. 2003;206:2049–57.

91. Martin F-PJ, Verdu EF, Wang Y, Dumas M-E, Yap IKS, Cloarec O, et al. Transgenomic metabolic interactions in a mouse disease model: interactions of *Trichinella spiralis* infection with dietary *Lactobacillus paracasei* supplementation. *J Proteome Res*. 2006;5:2185–93.
92. Wang Y, Holmes E, Nicholson JK, Cloarec O, Chollet J, Tanner M, et al. Metabonomic investigations in mice infected with *Schistosoma mansoni*: an approach for biomarker identification. *Proc Natl Acad Sci U S A*. 2004;101:12676–81.
93. Munger J, Bajad SU, Collier HA, Shenk T, Rabinowitz JD. Dynamics of the cellular metabolome during human cytomegalovirus infection. *PLoS Pathog*. 2006;2:e132.
94. Towler MC, Hardie DG. AMP-activated protein kinase in metabolic control and insulin signaling. *Circ Res*. 2007;100:328–41.
95. Lin J, Handschin C, Spiegelman BM. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab*. 2005;1:361–70.
96. Kim J, Tchernyshyov I, Semenza GL, Dang CV. HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab*. 2006;3:177–85.
97. Holness MJ, Sugden MC. Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation. *Biochem Soc Trans*. 2003;31(Pt 6):1143–51.
98. Doblado M, Moley KH. Facilitative glucose transporter 9, a unique hexose and urate transporter. *Am J Physiol Endocrinol Metab*. 2009;297:E831–5.
99. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
100. Andrews S: FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> 2010. Access 14 October 2015
101. Aronesty E: Command-line tools for processing biological sequencing data. *ea-utils* 2011.
102. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14:R36.
103. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511–5.
104. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
105. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
106. Cingolani P, Platts A, Wang LLL, Coon M, Nguyen T, Wang LLL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92.
107. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program. *SnpSift Front Genet*. 2012;3:35.
108. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31:46–53.
109. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562–78.
110. Wallace IM, O'Sullivan O, Higgins DG, Notredame C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res*. 2006;34:1692–9.
111. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5:113.
112. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059–66.
113. Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res*. 2005;15:330–40.
114. Pei J, Sadreyev R, Grishin NV. PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics*. 2003;19:427–8.
115. Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, et al. Fast statistical alignment. *PLoS Comput Biol*. 2009;5:e1000392.
116. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3.
117. Quang LS, Gascuel O, Lartillot N, Lirmm B, Cedex M: Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 2008;24(20):2317–3.
118. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. 2009;25:2286–8.
119. Cali A, Takvorian PM. Developmental Morphology and Life Cycles of Microsporidia. In: Weiss LM, Becnel JJ, editors. *Microsporidia Pathog Oppor*. 2014.
120. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010;27:221–4.
121. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57:289–300.
122. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 2014;42(Database issue):D199–205.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

